CERTIFICATE OF MAILING
37 C.F.R. 1.8

I hereby certify that this correspondence is being deposited with the U.S. Postal Service with sufficient postage as First Class Mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on the date below:

Oct. 4, 2006 — Date

~~Ann King~~ Eric B. Min

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:
Igor Dozmorov
Michael Centola

Serial No.: 10/693,656

Filed: October 24, 2003

For: AN ASSOCIATIVE ANALYSIS OF GENE EXPRESSION ARRAY DATA

Group Art Unit: 1631

Examiner: Marina I. Miller

Atty. Dkt. No.: OMRF:013US

DECLARATION UNDER 37 C.F.R. § 1.131

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

I, Igor Dozmorov, Ph.D., declare that:

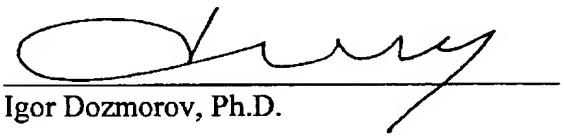
1. I am an inventor on the above-captioned application.

2. It is my understanding that the examiner in charge of the above-captioned application has advanced a rejection of claims 1, 3-5, 8-9, and 12 over the reference Xiao *et al.*, "Assessment of differential gene expression in human peripheral nerve injury," *BMC Genomics* 2002, 3:28 (27 September 2002) ("Xiao article"). I understand that the Xiao article lists a publication date of September 27, 2002.

3. The subject matter of the rejected claims was invented prior to the listed publication date of the Xiao article. As support of this statement, I have attached hereto the following paper: Dozmorov I, Centola M., "An associative analysis of gene expression array data." *Bioinformatics*, 19(2):204-11 (2003) ("Bioinformatics article," attached as Appendix A). The Bioinformatics article, which lists a "revised on" date of May 27, 2002, discloses the subject matter of the rejected claims. This demonstrates invention prior to the September 27, 2002 publication date of the Xiao article.

4. I hereby declare that all statements made herein of my knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

10/3/06
Date



Igor Dozmorov, Ph.D.

APPENDIX A

BIOINFORMATICS

Vol. 19 no. 2 2003
Pages 204–211

An associative analysis of gene expression array data

Igor Dozmorov and Michael Centola*

Department of Arthritis and Immunology, Oklahoma Medical Research Foundation, Oklahoma City, OK 73105, USA

Received on March 1, 2002; revised on May 27, 2002; accepted on July 11, 2002

ABSTRACT

Motivation: We face the absence of optimized standards to guide normalization, comparative analysis, and interpretation of data sets. One aspect of this is that current methods of statistical analysis do not adequately utilize the information inherent in the large data sets generated in a microarray experiment and require a tradeoff between detection sensitivity and specificity.

Results: We present a multistep procedure for analysis of mRNA expression data obtained from cDNA array methods. To identify and classify differentially expressed genes, results from standard paired *t*-test of normalized data are compared with those from a novel method, denoted an associative analysis. This method associates experimental gene expressions presented as residuals in regression analysis against control averaged expressions to a common standard—the family of similarly computed residuals for low variability genes derived from control experiments. By associating changes in expression of a given gene to a large family of equally expressed genes of the control group, this method utilizes the large data sets inherent in microarray experiments to increase both specificity and sensitivity. The overall procedure is illustrated by tabulation of genes whose expression differs significantly between Snell dwarf mice (dw/dw) and their phenotypically normal littermates (dw/+, +/+). Of the 2352 genes examined only 450–500 were expressed above the background levels observed in nonexpressed genes and of these 120 were established as differentially expressed in dwarf mice at a significance level that excludes appearance of false positive determinations.

Contact: igor-dozmorov@omrf.ouhsc.edu

INTRODUCTION

Analysis of the data from large-scale mRNA expression studies is nontrivial due to the complexity and size of data sets and the fact that technical variation can be introduced at different stages in array production and processing. Establishing well specified and carefully validated procedures for standardization and normalization of data sets from individual specimens is a key first step in analysis, but no single method has proven free from ambiguity. Selection criteria based on the ratio of measured expression levels fails to account for intra-group variations (i.e. normal biologic variance) and can result in false positive selections (Kerr *et al.*, 2000; Dozmorov *et al.*, 2002). More progressive statistical approaches such as regression analysis, multidimensional scaling, or principal component analysis, have been cogently criticized on a number of grounds, including the influence of outliers (i.e. genes expressed to different degrees among samples), on the parameters of linear regression, principal axis choice, and the absence of information about variability of individual expression levels within homogeneous groups of samples. Nonetheless, attempts at restricting the influence of outliers and noncorrelated weak signals has not resulted in the development of recognized standards (Newton *et al.*, 2001; Wu, 2001).

Additionally, current statistical methods do not adequately address the mutually exclusive characteristics of sensitivity and specificity. The common practice of using low thresholds for selection of significance ($P < 0.05$) can also result in a large number of false positive selections. This is especially problematic for high-density arrays as the number of false positive selections expected to occur by chance may limit the ability to perform higher order analyses, such as molecular pathway identification or disease subphenotyping, that require groups of differentially expressed genes to be accurately predicted. Attempts to increase stringency by raising the threshold of significance above this value can also be problematic as it will cause a compensatory decrease in sensitivity and resultant increase in false negative selections. The use of large numbers of replicates is able improve this situation (Glynne *et al.*, 2000), although it can be expensive and labor intensive. Herein we describe a novel statistical method of comparative analysis of cDNA array data. The method, denoted 'associative analysis', supplements the standard procedure of multiple paired comparisons by associating the expression level of each gene in the

*To whom correspondence should be addressed.

experimental group with a family of similarly and stably expressed genes in the control group. This associative analysis enhances the sensitivity of selections beyond previously described modifications of the *t*-test (Miller *et al.*, 2001) and increases the number of differentially expressed genes identified without significantly increasing the misidentification of false positives.

In our previous publications (Dozmorov *et al.*, 2001, 2002), some aspects of these normalization procedures have been applied to identification of differentially expressed genes in mice of the dwarf genotype (Ames—homozygous for the *Prop* 1^{df} mutation, and Snell—homozygous for the *Pit*1^{dw} mutation). Dwarf mice demonstrate similar deficiencies in pituitary dysfunction leading to decreased production of growth hormone, prolactin and thyroid-stimulating hormone and severe alterations in gene expression profiles relative to wild type mice (Pfaffle *et al.*, 1999). Herein, for the first time we apply the full suite of statistical procedures discussed above to these data sets and fully delineate the methods such that they can be assessed and employed by other groups.

SYSTEM AND METHODS

Experimental methods

This work uses the same raw experimental data and materials as a previous publication (Dozmorov *et al.*, 2002). This method included screening a commercial array of 2352 mouse genes with cDNA derived from Snell dwarf mice liver and their sibling controls. Eight 6 month old male homozygous Snell dwarf mice (dw/dw) and eight normal age and sex matched siblings (dw/+ or +/+) were studied. The raw data are available at http://www.omrf.ouhsc.edu/OMRF/Research/09/DozmorovI.asp.

Tissue collection, preparation of RNA and cDNA probes, and hybridization was done as described (Dozmorov *et al.*, 2002).

Outline of normalization and analysis procedures

1. Normalization of each expression profile to its own background, with selection of the genes expressed above background for subsequent adjustment and comparison. Note: the 'expressed' genes are selected as **not associated** with a representative homogeneous family of background level values having normal distribution (Figure 1).
2. Adjustment of the normalized profiles to each other by robust regression analysis of genes expressed above background. In this analysis potential outliers are identified and their contribution to the calculations down-weighted in an iterative manner, diminishing or excluding their influence (Figure 2). All expression profiles of both control and experimental groups are then re-scaled to a common standard—the averaged profile of the control group. An alternative procedure for outliers exclusion is based on the selection of equally expressed genes as homogeneous family of genes with normally distributed residuals measured as deviations from the regression line calculated against the averaged profile (Figure 3). Outliers are thereafter determined as having deviations **not associated** with this normal distribution presented by several hundred members.
3. Identification of a group of similarly expressed genes from control experiments, denoted 'reference group'(Figure 3), to be used for statistical analysis of differentially expressed genes using an associative *t*-test (below). The reference group is composed of a group of genes expressed above background levels with normal low variability of expression in control samples as determined by *f*-test, and whose residuals approximate a normal distribution, based on the Kolmogorov–Smirnov criterion.
4. Identification of genes differentially expressed in experimental vs control groups using three distinct statistical approaches (Figure 4). These analyses include:

— Selection of differentially expressed genes using a paired *t*-test (separate tests for a pair of replicates of each gene in the control and experimental groups) and the commonly accepted significance threshold of $P < 0.05$. A significant proportion of the genes identified as differentially expressed will be false positive determinations at this threshold level.

— A *t*-test using a Bonferroni correction for the significance threshold that effectively eliminates false positive determinations with simultaneous loss of the sensitivity, resulting in increased proportion of false negative determinations.

— An **associative** *t*-test in which the replicated residuals for each gene of the experimental group are compared with the entire set of residuals from the reference group defined above. The null hypothesis is checked to determine if gene expression in the experimental group is **associated** with the reference group defined in step 3. The significance threshold is corrected to make improbable the appearance of false positive determinations.

— Comparison of the selections from the paired *t*-test and associative *t*-tests to classify the differentially expressed genes identified as: (a) likely false positives (these are genes selected as differentially expressed by the paired *t*-test with $P < 0.05$, but not by the associative *t*-test); (b) real positives (selected

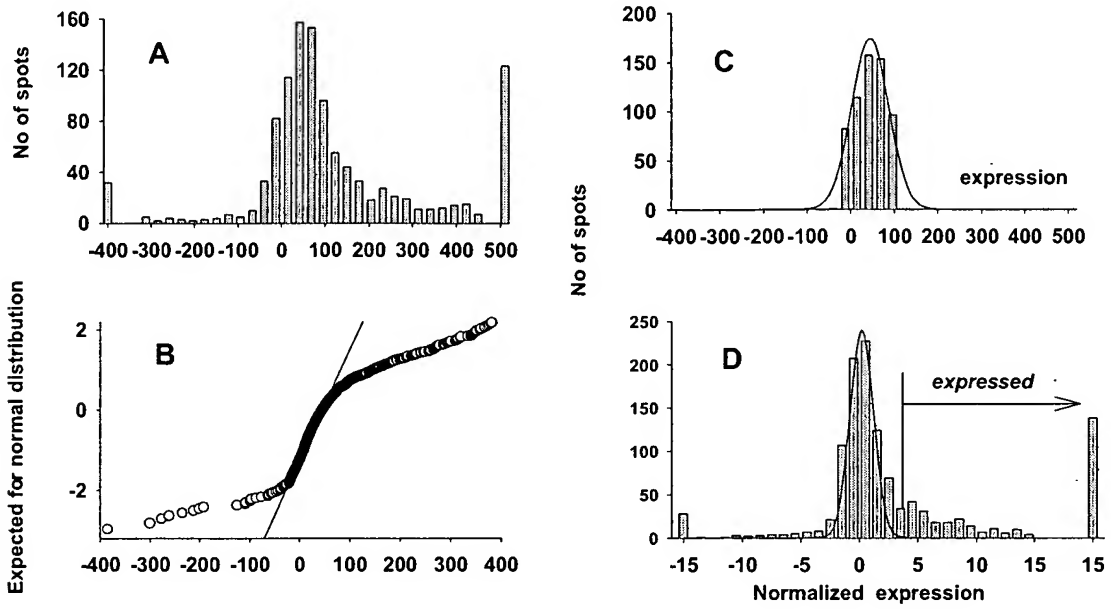


Fig. 1. Normalization of the gene expression profile to its own background. (a) A histogram showing the expression levels of the 1176 cDNA targets derived from the liver of normal control mice. These values conform poorly to a normal distribution, with extended upper and lower tails apparent. Values in the lower tail result from the background correction procedure (typically negative). Values in the upper tail correspond to genes expressed above background in a given sample. (b) Plot of the values with the proposed normal distribution Vs the real levels of expression. The straight line is a regression line for the central part of the plot—predominantly background noise. To identify the parameters of normal distribution for background, data are sorted in ascending order and, as a first approximation, the mean and SD estimates are computed for all spots. Spots at the high end and at the low end of this distribution are then discarded one by one in alternating manner if they exceed a criterion set 2 SDs beyond the mean of the remainder of the distribution. The resulting set of nondiscarded points (typically between 500 and 600 of the initial set of 1176) represents the fragment of normally distributed background values. This fragment is then used for the accurate estimation of the parameters of the normal distribution for background using a standard minimization procedure (Figure 1c). The mean and SD of normally distributed background spots are used for the raw intensity S normalization as $S' = (S - - - Av)/\mathrm{SD}$. The distribution of S' (1D) has mean of zero and SD = 1 over the set of background genes. The curve shows the distribution of these nonexpressed genes. The threshold 3SD = 3 was used for selection of genes expressed above background.

in both tests) (c) potential positives (genes selected in the associative test only).

RESULTS

Comparative analysis of gene expressions in the experimental group is begun by applying the procedures of normalization to background, and rescaling described above. Averaged data from the control group is used as a standard for data rescaling. The adjustment of data from the experimental group to averaged control data will produce the same order residuals for equally expressed genes and highlight the genes with extreme expression deviations (Figure 3c).

Single gene comparisons—paired t-test The paired t-test evaluates the difference between the means of each single gene expression in two groups employing the variance within groups as an error term. The use of the usual threshold $P = 0.05$ for the selection of differentially expressed genes will result in a significant proportion of false positive selections from experiments with thousands of elements, as is the case in array experiments. When using the Atlas 1.2 array set, about 50 false positive selections can be expected at this threshold if all genes are analyzed. This number can be substantially decreased if the analysis excludes genes that are not expressed in both groups. Approximately 250 genes were determined to be expressed in the experiments described herein. The proportion of false positive determinations expected in this group at $P = 0.05$, which is 12 to 13, still represents a significant portion of the total number of differentially expressed genes identified. Use of replicates results in a decrease of the proportion of false negative determinations though the proportion of false positives

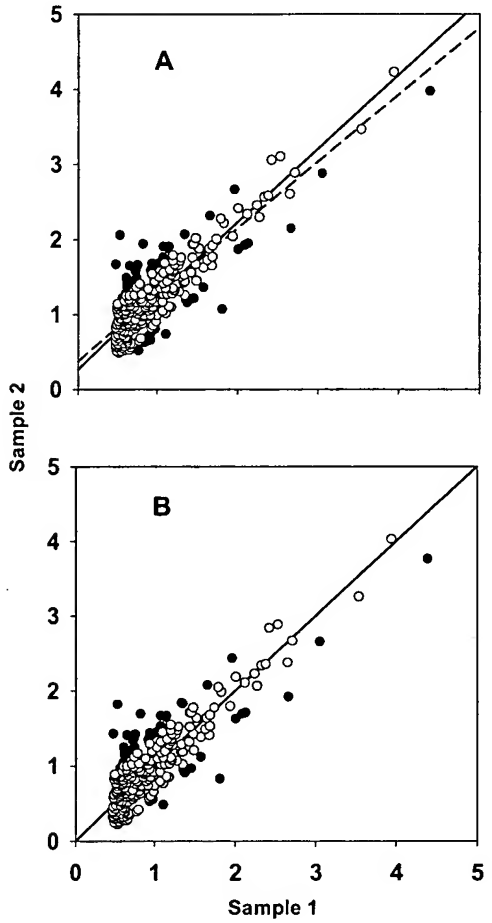


Fig. 2. Comparison of liver samples of two normal mice (Atlas I arrays as in Figure 1c). Each data set (S_1 and S_2) has been normalized with respect to its own set of background genes, as explained above. The values are shown on a logarithmic scale, and only 'expressed above background' values where S > Log (3) are included. Differentially expressed genes can be identified as those whose ratio of expression in two control samples does not fall on or close to the line describing similarly expressed genes (filled circles in Figure 2a). These genes denoted as 'outliers' were excluded from rescaling by use of a robust regression procedure in which the influence of outliers is down-weighted in a series of regression procedures (software—Number Cruncher Statistical System, Utah, 2001) with an influence function based on the use of least absolute deviations and with twenty subsequent cycles of the regression parameters estimations. The resulting plot for completely adjusted distributions is presented in Figure 2b with the final regression line passing through the origin with the slope equal to 45°.

remains relatively stable—around one third of all positive selections (Figure 4a and b). It is possible to decrease this proportion through the use of a corrected P-value.

Single gene comparison Bonferroni t-test The Bonferroni correction is the most common method employed to reduce the proportion of false positive determinations in multiple comparison analysis and it has been applied to array data (Miller *et al.*, 2001). In this method the stringency of the threshold P is increased to 0.05/(the number of compared values). For the expressed genes identified above $P = 0.05/250 = 2 \times 10^{-4}$. This increased threshold produces a new selection of differentially expressed genes with the complete absence of false positive determinations (Figure 4c). While specificity is increased in this analysis, sensitivity is sacrificed and a large number of false negatives, type II errors, are obtained. All selections obtained with Bonferroni t-test are present also within selections made in **Associative comparison** (see below).

Associative comparison It is possible to substitute the typical paired comparison of gene expressions between control and experimental groups with the comparison of their residuals. In this analysis it is determined if a given gene of the experimental group belongs to (or can be associated with) the reference group (as defined in **Experimental Methods**, Outline step 3). Denoted an associative t-test, it is actually a standard Student t-test applied to the comparison of expression deviations. An associative t-test dramatically increases the power of comparisons relative to a paired t-test. In the data analyzed herein, this is due to the fact that eight replicates from the control group are compared with several hundred values of the reference group. As a result a large number of positive determinations can be obtained with stringent thresholds (Figure 4d).

By comparing the results of these two tests, paired t-test with threshold $P < 0.05$, and associative t-test with threshold $P < 0.005 (P < 1/n$, where $n =$ number of genes analyzed from the experimental group), differentially expressed genes can be classified into three groups. Genes defined as differentially expressed by the paired t-test but not by the associative t-test are likely false positives. Genes identified in both analyses are likely real positives, that also include the small sub-group of genes selected by the Bonferroni t-test. Genes identified in the associative analyses are potentially real positives that require additional replicates to confirm.

We have used this analysis to identify genes that are differentially expressed between normal and dwarf mice and found 46 genes overexpressed in Snell dwarf mice; 49 genes expressed only in Snell mice; 12 genes overexpressed in normal control mice; 13 genes expressed only in normal mice. (Tables S1 A–D in the Supplementary section http://www.omrf.ouhsc.edu/OMRF/Research/09/Dozmorovl.asp). Of these selected genes 71 are previously reported as differentially expressed in Snell dwarf mice, associated with dwarfism, or strongly

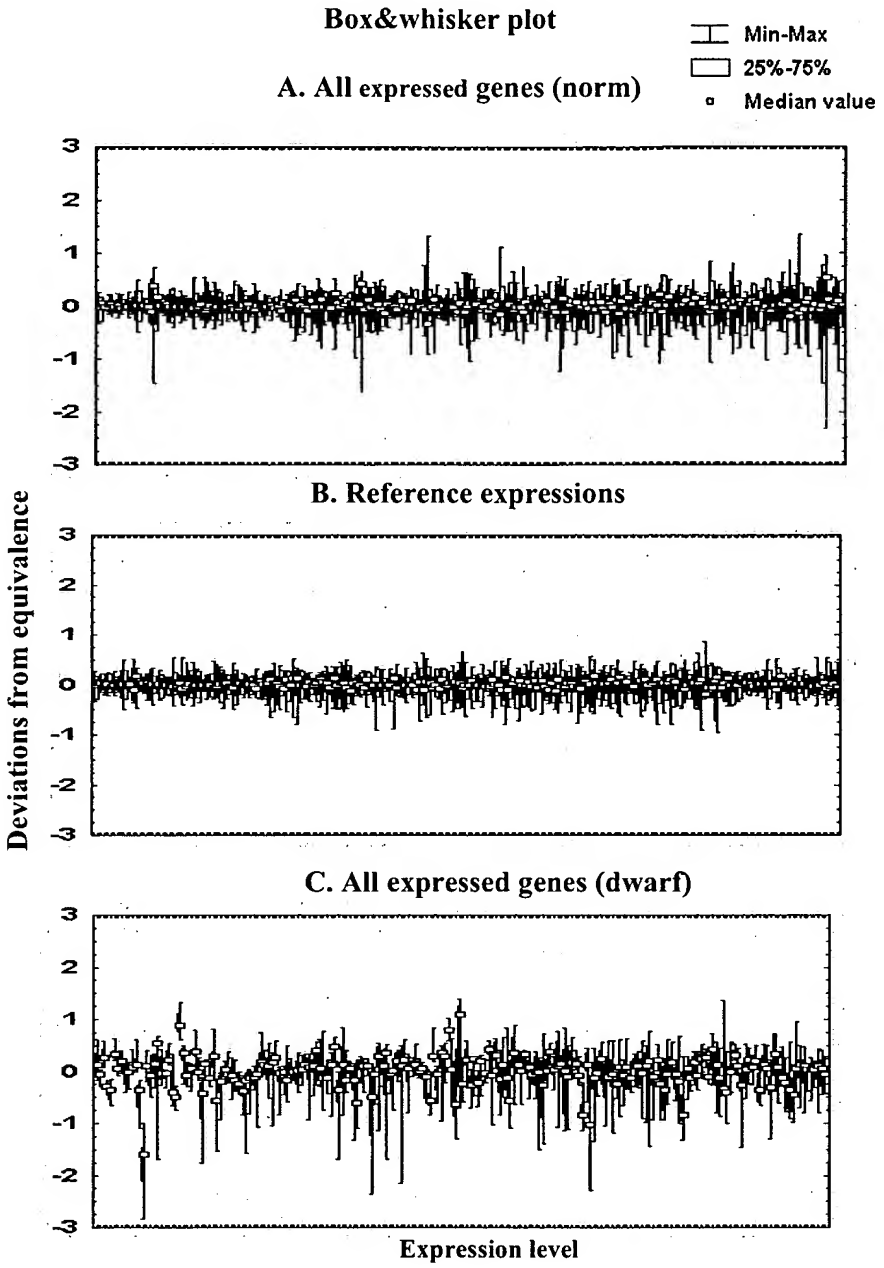


Fig. 3. Deviations of gene expression after rescaling to the averaged data in normal mice group (8 mice). (a) Variability of genes within the homogeneous control group. (b) The same data after exclusion of hypervariable genes with a SD statistically higher than the homogeneous control group (based on an F-criterion). (c) Deviation from normal control averages of gene expressions in dwarf mice samples.

associated with a similar hormonal status. An additional 10 selections were obtained by the new method and not obtained by previous analysis (Dozmorov *et al.*, 2002), whose relevance to dwarfism or similar hormonal status are supported by the indicated references (Table 2S). In addition, this new method was able to more correctly predict the expression levels of 11 genes verified by RT-PCR in the previous publication (Dozmorov *et al.*, 2002).

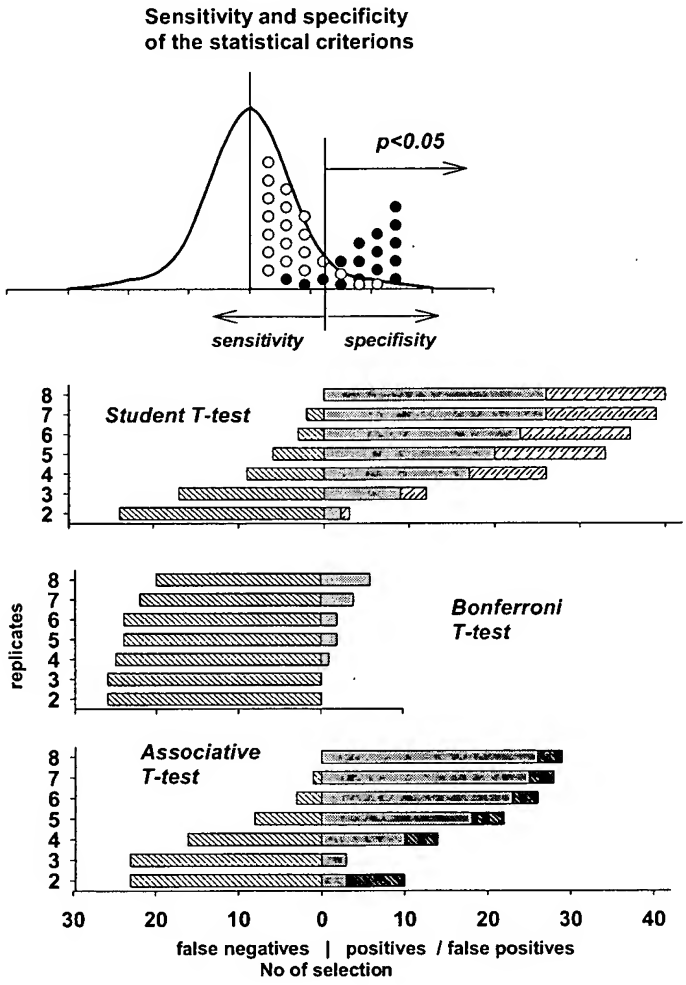


Fig. 4. Sensitivity and specificity of statistical comparison. Numbers of genes with statistically different expression in dwarf mice compared with their normal siblings were selected from 256 expressed genes presented on Atlas-I membrane using three different criterions—paired t-test ($P < 0.05$), Bonferroni t-test ($P < 0.05/256$), and associative t-test ($P < 0.0025$). Positive, false positive and false negative selections are shown with different filling as indicated.

DISCUSSION

We describe herein a useful and practical multistep procedure to analyze gene expression data from a cDNA array. The method is novel in that it: provides a robust means of normalizing one channel data using an internal standard; establishes a more precise procedure for data scaling by reducing the influence of outliers upon calculation of scalars; increases the sensitivity of differential gene identification without loss of specificity; and allows differentially expressed genes to be classified into distinct groups of probabilistically known or suspected differential expression.

We demonstrate here an opportunity to increase the power of statistical analysis using representative standards for selection of potential outliers. This general procedure is done three times in these analyses. The first representative standard is the family of genes whose hybridization signals are at or below the background level. Outliers from this standard are defined as 'expressed genes'. The second representative standard is the family of normally distributed residuals of equally expressed genes of the control group. Outliers from this group are hypervariable and differentially expressed genes that must be excluded from regression analysis for proper adjustment of pairs

of profiles under comparison. The third representative standard is the family of genes with low variability within replicate control samples. There are two types of outliers from this standard—hypervariable genes of the control group (which were excluded to create this standard) and differentially expressed genes of the experimental group—whose identification is the main goal of these analyses.

The necessity to initially exclude from comparisons expressed from nonexpressed genes was demonstrated herein with data obtained from Snell mice using Clontech Atlas arrays in which 600 genes were spotted in duplicate. Since two independent signals are measured for each gene on the membrane, the variation in intensity between the duplicated spots for a given gene can be used to assess signal reproducibility. If variation were independent of signal intensity the ratio of variation between duplicate spots would be distributed around 1 with small random variations. However, this was not observed for genes expressed below some threshold signal intensity. It is of note that this threshold corresponds to our determination of background. It is not clear if this so-called background threshold is due to technical limitations of measuring signal intensity on the array or if it is a real biologic threshold defined by genes that are not expressed. Operationally however, the addition of this exclusion criterion provides a logical cutoff between noncorrelative and correlative data and therefore improves the reliability of the comparative analysis carried on after this step. A similar conclusion has been presented (Newton *et al.*, 2001; Khan *et al.*, 2001) though with use of arbitrary exclusion criteria. While these exclusion criteria improve the homogeneity of selections made using ratios, the arbitrariness can be associated with loss of useful information about low abundance genes that can play an important role in regulatory biological processes.

We accomplish further enrichment of reliability utilizing and extending previously published theories on signal variation. There are different sources for fluctuations in residuals. Technological variations represent a random component of deviation and are therefore common for all expressions. Some publications demonstrate the dependence of technological fluctuations on the level of gene expression, and a resultant nonnormal distribution of these values (Kerr *et al.*, 2000; Newton *et al.*, 2001). The two main sources of heterogeneity in gene expression variations are indicated in Rocke and Durbin (2001) as, the 'additive component', prominent at low expression levels, and the 'multiplicative component', prominent at high expression levels. The intensity measurement $y_{i,j}$ for gene $i \in I = \{i_1, \ldots, i_n\}$ in sample $j \in J = \{j_1, \ldots, j_m\}$ is modeled by the equation (Rocke and Durbin, 2001; Zien *et al.*, 2001) $y_{i,j} = \alpha_{i,j} + \mu_{i,j} \times e^{\eta} + \varepsilon_{i,j}$ where α—is the normal background (and independent of expression level), μ—the expression level in arbitrary units, ε—is first error term (additive)—which represents the standard deviation of background, and η—is the second error term, which represents the proportional error (multiplicative). The first error term is excluded in our analysis by eliminating expression values at or below background levels. The second error term is transformed from multiplicative (and therefore expression-dependent, increasing in proportion to expression level (Lee *et al.*, 2000), into additive—or expression independent) by log-transformation of data as in Rocke and Durbin (2001): $\log(y) = \log(\mu) + \eta$, where η is the residual for log-transformed data. The independence of η from individual gene expressions is proven by the vendor (Atlas Manual, 2000) and confirmed with the Kolmogorov–Smirnov normality test in our experiments.

Not surprisingly, we have found that the number of repetitions is critical in achieving adequate specificity (low false positives) and sensitivity (low false negatives). Due to the stochastic character of the above-mentioned fluctuations, replication and averaging is a sensible method to reduce the noise level. Only those transcripts that are truly altered by an experimental factor will have a reproducible change and become more statistically significant with repetition; those changes that result from noise will not become more significant with repetition. Thus, sensitivity increases with repetition at a fixed specificity. Both the paired t-test and the associative t-test demonstrate similar improvement in sensitivity through replication. However, the specificity of paired t-test remains unchanged when using from 4 to 8 replicates. This is due to the use of the necessity to use conservative methods to protect from false positive determinations when using the paired t-test. These methods result in the loss of information about the majority of false negative expression differences. This information once lost, is not regained through additional replicates. In the associative t-test selections are made at a significance threshold high enough to exclude the appearance of false positive determinations. However, the number of comparisons made between a given experimental gene and the family of similarly expressed genes in the control condition assures that few false negative determinations will occur. Increased repetition can therefore be used to enhance the overall statistical significance of the selections made using this method (Figure 4). Conformation of the increased sensitivity of this method was obtained from a literature search of genes whose expression has been shown to be different in Snell mice and related model systems . At the level of sensitivity with less than one false positive determination the associative method selects a larger number of differentially expressed genes documented in the literature to have links with dwarfism or similar abnormalities in hormonal status than previous methods

utilizing a paired analysis. Only half (around 30) of the differences obtained by microarray studies that utilized a standard paired analysis (Dozmorov *et al.*, 2002) were confirmed in the current analysis. Importantly, only those genes confirmed by the associative method have been shown to be related with a premature aging phenotype in empirical studies, suggesting the methods described herein do indeed increase the specificity of differential gene identification.

The associative method also enhances the information obtained from microarray experiments beyond common approaches because it discriminates between genes that are differentially expressed from those that are expressed only in one state. For example Calgranulin B has been shown previously by RT-PCR to be undetectable in normal mice, as predicted by the method described herein, yet selected as differentially expressed in a previous analysis utilizing only a standard paired comparison (Dozmorov *et al.*, 2002). Procedures similar to associative analysis have been previously proposed (Newton *et al.*, 2001; Rocke and Durbin, 2001). However, there are critical differences between these methods and ours. For example in Rocke and Durbin (2001) all genes were used as a reference group without excluding hypervariable genes. The presence of hypervariable genes increases the standard deviation of the reference group thus reducing the power of the associative analysis. Moreover, inclusion of hypervariable genes results in a nonnormally distributed reference group preventing the use of common statistical tests.

By testing the hypothesis of association of any potential outlier with a large representative standard (typically several hundred elements) the statistical power of the analysis is increased over that achieved with traditional single gene comparisons, which are powered only by the numbers of replicates. The higher power of the associative test, thus, increases sensitivity without loss of specificity. When used in combination with a traditional paired analysis, this increased statistical power also allows the use of traditional low level significance cutoffs in the standard paired analysis ($P < 0.05$) without the risk of including false positive selections. The associative analysis is therefore based on an idea opposite to the commonly held view that large-scale array experiments suffer from compensatory tradeoffs in sensitivity and specificity. In fact the procedures presented herein demonstrate that large-scale data sets are extraordinarity information rich and provide a means for discriminating common technical variation from individual biological variability.

ACKNOWLEDGEMENTS

This work was supported by grants from the National Institutes of Health National Center for Research Resources (grant numbers NIH 1 P20 RR15577 and NIH 1P20 RR16478), the Nathan Shock Center, and The Fund for Arthritis and Inflammatory Research (FAIR). We acknowledge Drs Richard Miller and Ivan Lefkovits for stimulating discussions regarding this material and Andrzej Bartke, Igor Sidorov, Nicholas Knowlton for their evaluations of the manuscript.

REFERENCES

Atlas Manual (2000) Clontech Atlas cDNA expression arrays user manual. PT3140-1.

Dozmorov,I., Bartke,A. and Miller,R.A. (2001) Array-based expression analysis of mouse liver genes: effect of age and of the longevity mutant $Prop1^{df}$. *J. Gerontol.*, **56A**, B72–B80.

Dozmorov,I., Galecki,A., Chang,Y., Krzesicki,R., Vergara,M. and Miller,R.A. (2002) Gene expression profile of long-lived Snell dwarf mice. *J. Gerontol.*, **57A**, B1–B10.

Glynne,R.J., Ghandour,G. and Goodnow,C.C. (2000) Genomic-scale gene expression analysis of lymphocyte growth, tolerance and malignancy. *Curr. Opin. Immunol.*, **12**, 210–214.

Kerr,M.K., Martin,M. and Churchill,G.A. (2000) Analysis of variance for gene expression microarray data. *J. Comput. Biol.*, **7**, 819–837.

Khan,J., Wei,J.S., Ringner,M., Saal,L.H., Ladanyi,M., Westermann,F., Berthold,F., Schwab,M., Antonescu,C.R., Peterson,C. and Meltzer,P.S. (2001) Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature. Med.*, **7**, 673–679.

Lee,M.L., Kuo,F.C., Whitmore,G.A. and Sklar,J. (2000) Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc. Natl Acad. Sci. USA*, **97**, 9834–9839.

Miller,R.A., Galecki,A. and Shmookler-Reis,R.J. (2001) Interpretation, design, and analysis of gene array expression experiments. *J. Gerontol.*, **56A**, B52–B57.

Newton,M.A., Kendziorski,C.M., Richmond,C.S., Blattner,F.R. and Tsui,K.W. (2001) On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data. *J. Comput. Biol.*, **8**, 37–52.

Pfaffle,R.W., Blankenstein,O., Wuller,S. and Kentrup,H. (1999) Combined pituitary hormone deficiency: role of Pit-1 and Prop-1. *Acta. Paediatr.*, **Suppl 88**, 33–41.

Rocke,D.M. and Durbin,B.A. (2001) A model for measurement error for gene expression analysis. *J. Comput. Biol.*, **8**, 557–569.

Wittes,J. and Friedman,H.P. (1999) Searching for evidence of altered gene expression: a comment on statistical analysis of microarray data. *J. Nat. Cancer Inst.*, **91**, 400–401.

Wu,T.D. (2001) Analysing gene expression data from DNA microarrays to identify candidate genes. *J. Pathol.*, **195**, 53–65.

Yang,M.C., Ruan,Q.G., Yang,J.J., Eckenrode,S., Wu,S., McIndoe,R.A. and She,J.X. (2001) A statistical method for flagging weak spots improves normalization and ratio estimates in microarrays. *Physiol. Genomics.*, **7**, 45–53.

Zien,A., Aigner,T., Zimmer,R. and Lengauer,T. (2001) Centralization: a new method for the normalization of gene expression data. *Bioinformatics*, **1**, 1–9.